

VU Research Portal

De cognitieve interview methode: Bespreking van een alternatieve manier om testbias te onderzoeken

Oostrom, J.K.; Born, M.Ph.

published in

Gedrag en Organisatie
2013

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Oostrom, J. K., & Born, M. P. (2013). De cognitieve interview methode: Bespreking van een alternatieve manier om testbias te onderzoeken. *Gedrag en Organisatie*, 26, 205-222.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

REEKS 'WERVING EN SELECTIE'

De cognitieve interviewmethode

Bespreking van een alternatieve manier om testbias te onderzoeken

Janneke K. Oostrom & Marise Ph. Born*

Aangezien selectie-instrumenten dikwijls etnische scoreverschillen genereren, is het creëren van een divers personeelsbestand voor veel organisaties een complexe uitdaging. In de ontwikkeling van selectie-instrumenten moet daarom worden getracht om onrechtvaardige scoreverschillen tussen etnische groepen zo veel mogelijk te reduceren. In dit artikel wordt een methode besproken die onderzoekers en testontwikkelaars hierbij kan helpen, namelijk de cognitieve interviewmethode. Deze kwalitatieve methode voor het evalueren van vragenlijsten wordt reeds toegepast in de gezondheidszorg. In dit artikel wordt beargumenteerd dat deze methode ook nuttig kan zijn op het gebied van personeelsselectie. De cognitieve interviewmethode kan gebruikt worden om te onderzoeken of items inhoudelijk dezelfde betekenis toegeschreven krijgen door respondenten met verschillende etnische achtergronden.

1 Inleiding

De selectiepraktijk is sterk onderhevig aan sociaaleconomische en maatschappelijke veranderingen. Een van deze veranderingen, die een duidelijke uitwerking heeft op de selectiepraktijk, is de toenemende diversiteit op de arbeidsmarkt (Derous, Van der Velde & Born, 2011). Het selecteren van een etnisch divers personeelsbestand brengt complexe uitdagingen met zich mee. Selectie-instrumenten zijn primair ontwikkeld en gevalideerd voor de autochtone bevolkingsgroep (Derous et al., 2011) en genereren dikwijls scoreverschillen tussen de autochtone groep en etnische minderheidsgroepen (Ployhart & Holtz, 2008). De omvang van de scoreverschillen bepaalt of een selectie-instrument 'adverse impact' vertoont. Volgens de definitie van Ironson, Giuon en Ostrander (1982) is er sprake van 'adverse impact' wanneer de selectieratio van de meerderheidsgroep substantieel groter is dan de selectieratio van de minderheidsgroep. Naast het feit dat 'adverse impact' het creëren van een divers personeelsbestand hindert, stelt ook wet- en

* Janneke K. Oostrom is verbonden aan de Vrije Universiteit Amsterdam, Afdeling Sociale en Organisationspsychologie. Correspondentieadres: Vrije Universiteit Amsterdam, Faculteit der Psychologie en Pedagogiek, Afdeling Sociale en Organisationspsychologie, 1B-25, Van der Boechorststraat 1, 1081 BT Amsterdam, tel. 020-5983382, e-mail j.k.oostrom@vu.nl. Marise Ph. Born is verbonden aan de Erasmus Universiteit Rotterdam, Instituut voor Psychologie, Sectie Arbeids- en Organisationspsychologie.

regelgeving op het terrein van gelijke behandeling (bijv. de Algemene Wet Gelijke Behandeling) dat het gebruik van instrumenten en procedures die personen, behorend tot etnische minderheidsgroepen, onevenredig en zonder objectieve rechtvaardiging benadelen, vermeden dient te worden. In de ontwikkeling van selectie-instrumenten moet om deze reden getracht worden de scoreverschillen tussen de autochtone groep en de etnische minderheidsgroepen in te perken.

De doelstelling van het huidige artikel is de beschrijving en kritische bespreking van de cognitieve interviewmethode als een mogelijke manier om onrechtvaardige scoreverschillen van diverse soorten selectie-instrumenten in te perken. Met onrechtvaardige scoreverschillen worden scoreverschillen bedoeld die ontstaan ten gevolge van kennis, eigenschappen en vaardigheden die niet relevant zijn voor de uitoefening van de toekomstige functie en bijgevolg ontstaan zijn door irrelevante testvariantie. De cognitieve interviewmethode geeft inzicht in de conceptuele equivalentie van instrumenten (Barofsky, 2000). Met conceptuele equivalentie wordt bedoeld dat informatie zoals weergegeven in items of instructieteksten op dezelfde manier geïnterpreteerd worden door leden van verschillende groepen. Op basis van de uitkomsten van de cognitieve interviewmethode kunnen items of instructieteksten gereviseerd worden of nieuwe items ontwikkeld worden zodat deze geschikt zijn voor gebruik in verschillende etnische groepen (Nápoles-Springer, Santoyo-Olsson, O'Brien & Stewart, 2006). In de volgende paragrafen zal een overzicht worden gegeven van de huidige methoden om bias te reduceren. Vervolgens zal worden uitgelegd wat de cognitieve interviewmethode inhoudt en hoe deze methode een bijdrage zou kunnen leveren aan de ontwikkeling van selectie-instrumenten die gebruikt kunnen worden voor het selecteren van een divers personeelsbestand. Tot slot zal er een kritische reflectie op deze nieuwe methode worden gegeven.

2 Methoden om bias te reduceren

Selectie-instrumenten genereren dikwijls scoreverschillen tussen de autochtone groep en etnische minderheidsgroepen (Ployhart & Holtz, 2008). Zo laten cognitieve vaardigheidstests, vaak valide voorspellers van werkprestaties (Schmidt & Hunter, 1998), een scoreverschil van ongeveer één standaarddeviatie (*SD*) zien (bijv. Roth, Bevier, Bobko, Switzer & Tyler, 2001). Maar ook andere veelgebruikte selectie-instrumenten vertonen scoreverschillen in het nadeel van etnische minderheidsgroepen, bijvoorbeeld voor het gestructureerde interview is het scoreverschil $0.23\ SD$ (Bobko, Roth & Potosky, 1999) en voor de situationele beoordelingstest is het scoreverschil $0.40\ SD$ (Nguyen, McDaniel & Whetzel, 2005). De literatuur over etnische scoreverschillen biedt verschillende verklaringen voor deze groepsverschillen, waaronder omgevingsgerelateerde factoren (bijv. Ceci & Williams, 1997), genetische en biologische factoren (bijv. Jensen, 1981; Rushton, 1995) en testbias (bijv. Schmitt, 1989; Van de Vijver & Tanzer, 2004). Het huidige artikel richt zich op testbias.

Er is sprake van testbias wanneer de verschillen in geobserveerde scores niet overeenkomen met de verschillen in het concept waarop groepen worden vergele-

ken (Poortinga, 1989). Er kunnen drie typen testbias worden onderscheiden, namelijk itembias, methodebias en conceptbias (Van de Vijver & Tanzer, 2004). Er is sprake van itembias wanneer bepaalde items een verschillende betekenis hebben voor leden van verschillende etnische groepen (Holland & Wainer, 1993). Methodebias is bias die veroorzaakt wordt door kenmerken van de test, zoals de antwoordschaal of de persoon die de test afneemt (Poortinga, 2001). Er is sprake van conceptbias wanneer het concept dat gemeten wordt met de test, niet equivalent is voor verschillende etnische groepen. Voor een meer gedetailleerd overzicht van de drie typen testbias en de mogelijke bronnen verwijzen wij naar Van de Vijver en Tanzer (2004).

Er zijn reeds veel strategieën voorgesteld om testbias te reduceren. Deze kunnen geclusterd worden in twee typen, namelijk op de persoon gerichte strategieën en op het instrument gerichte strategieën. Tabel 1 biedt een overzicht van de verschillende strategieën, het effect met betrekking tot het daadwerkelijk reduceren van de scoreverschillen en enkele sleutelpublicaties.

Zoals uit tabel 1 kan worden afgelezen, blijken de strategieën niet altijd effectief te zijn in het daadwerkelijk reduceren van etnische scoreverschillen (zie ook Born, 2010; De Corte & Lievens, 2003). Op het instrument gerichte strategieën zijn in het algemeen effectiever in het reduceren van etnische scoreverschillen dan op de persoon gerichte strategieën. Veel van deze op het instrument gerichte strategieën doen echter afbreuk aan de voorspellende waarde van het selectie-instrument (Ployhart & Holtz, 2008). Dit is bijvoorbeeld het geval bij 'score banding' (strategie 10; zie bijv. Campion et al., 2006). Een effectieve strategie die geen afbreuk doet aan de voorspellende waarde van een selectie-instrument, is het meten van een grotere verscheidenheid in kennis, eigenschappen en vaardigheden. Ook het minimaliseren van de benodigde taalkennis tot het niveau dat vereist wordt in de toekomstige baan, is effectief in het doen reduceren van etnische scoreverschillen en doet in de meeste gevallen geen afbreuk aan de validiteit van het instrument (bijv. Arthur, Edwards & Barrett, 2002). Lievens en Sackett (2006) toonden aan dat het gebruik van videoclipen in plaats van geschreven teksten de voorspellende waarde van een SJT zelfs kan laten toenemen.

Geen enkele strategie is echter in alle situaties even effectief of praktisch toepasbaar. De zoektocht naar effectieve strategieën die geen afbreuk doen aan de voorspellende waarde van het instrument, blijft continueren. In het huidige artikel bespreken we een strategie die regelmatig gebruikt wordt voor het onderzoeken en reduceren van testbias in de gezondheidszorg, maar nog relatief onbekend is in de selectiepsychologie. Deze methode, genaamd de cognitieve interviewmethode, kan testontwikkelaars mogelijk helpen bij het inperken van onrechtvaardige scoreverschillen op diverse soorten selectie-instrumenten.

Tabel 1 *Overzicht van strategieën om etnische scoreverschillen op selectietests te reduceren*

| Strategie | Effect met betrekking tot reduceren scoreverschillen | Belangrijke referenties |
|---|--|---|
| Op de persoon gerichte strategieën | | |
| 1. Hertesten. Sollicitanten de kans geven om de test opnieuw te laten maken. | Klein | Schleicher, Van Iddekinge, Morgeson en Campion (2010) Sin, Farr, Murphy en Hausknecht (2004) |
| 2. Testtraining of meerdere oefenitems aanbieden. Sollicitanten van tevoren laten oefenen. | Klein en inconsistent | Sackett, Schmitt, Ellingson en Kabin (2001) Tanzer, Gittler en Ellis (1995) |
| 3. Gericht rekruteren. Het verhogen en behouden van etnische minderheden in de sollicitantenpool. | Klein | Ryan, Sacco, McFarland en Kriska (2000) Tam, Murphy en Lyall (2004) |
| 4. Kandidaatreacties verbeteren. Door reacties van leden minderheidsgroep te verbeteren wordt motivatie verhoogd. | Klein | Hausknecht, Day en Thomas (2004) Ryan (2001) |
| 5. Antwoordstijlen in kaart brengen. Controleren voor verschillen in sociale wenselijkheid. | Klein | Hamamura, Heine en Paulhus (2008) Hui en Triandis (1989) |
| Op het instrument gerichte strategieën | | |
| 6. Alternatieve instrumenten gebruiken. Een grotere verscheidenheid in kennis, eigenschappen en vaardigheden meten. | Groot | Hough, Oswald en Ployhart (2001) Schmitt, Clause en Pulakos (1996) |
| 7. Proxies gebruiken. Intelligentie meten aan de hand van schoolcijfers. | Klein tot gemiddeld | Berry, Gruys en Sackett (2006) Roth en Bobko (2000) |
| 8. Specifieke, smallere metingen gebruiken. Schalen in plaats van gehele instrumenten gebruiken. | Klein tot gemiddeld | Hough et al. (2001) |
| 9. Verscheidenheid in kennis, eigenschappen en vaardigheden vergroten. Gebruiken van cognitieve en non-cognitieve instrumenten. | Groot | Bobko et al. (1999) Ryan, Ployhart en Friedel (1998) |

Tabel 1
(Vervolg)

| Strategie | Effect met betrekking tot reduceren scoreverschillen | Belangrijke referenties |
|--|--|---|
| 10. 'Score banding'. Scores onderverdelen in banden waarbinnen geen onderscheid wordt gemaakt. | Groot indien binnen de banden voorkeur wordt gegeven aan leden van etnische minderheidsgroep | Aguinis (2004) Sackett en Roth (1991) |
| 11. Gewichten toekennen. Meer gewicht toekennen aan instrumenten of criteria die minder scoreverschillen genereren. | Klein tot gemiddeld | De Corte, Lievens en Sackett (2007) Hattrup, Rock en Scalia (1997) |
| 12. Minimaliseren van benodigde taalkennis. Benodigde taalkennis verlagen tot niveau van de baan. | Groot | Arthur et al. (2002) Chan en Schmitt (1997) |
| 13. Items gebruiken die geen inhoudelijke betekenis hebben. Alle groepen even (on)bekend laten zijn met de items. | Klein | DeShon, Smith, Chan en Schmitt (1998) Whitney en Schmitt (1997) |
| 14. Psychometrische procedures (bijv. differential item functioning [DIF]). Items identificeren en verwijderen van problematische items. | Klein | Hough et al. (2001) Sackett et al. (2001) Voddegel Matzen, Van der Molen en Dudink (1994) |
| 15. Review panels. Panel laten bekijken of items gebruikt kunnen worden in alle groepen. | Geen data beschikbaar | Grill en Bariel (1977) Reckase (1996) |
| 16. Tijdslimiet verwijderen. | Geen tot klein | Sackett et al. (2001) |
| 17. Trainen van testafnemers (bijv. interculturele sensitiviteit) en uitvoerige scoringsinstructies aanbieden. | Klein tot gemiddeld, vooral effectief bij selectie-interviews | Gittler (1990) Gutjahr (1985) |
| 18. Controleren voor contextvariabelen (bijv. opleidingsniveau, arousal). | Gemiddeld | Poortinga en Van de Vijver (1987) |

* Deze tabel is grotendeels gebaseerd op tabel 2 uit Ployhart en Holtz (2008) en tabel 3 uit Van de Vijver en Tanzer (2004).

3 De cognitieve interviewmethode

De cognitieve interviewmethode is een kwalitatieve methode die inzicht biedt in de cognitieve processen die in werking treden bij het lezen van items of teksten. Respondenten krijgen een instrument (bijv. een persoonlijkheidsvragenlijst) voorgelegd en worden tegelijkertijd gevraagd om alle gedachten die opkomen tijdens het lezen van de teksten of items hardop uit te spreken (Beatty & Willis, 2007). Deze informatie wordt gebruikt om te bepalen of de teksten of items de gedachten oproepen die de ontwikkelaars van het instrument ook daadwerkelijk wensten op te roepen en indien nodig de items of teksten te verwijderen of te verbeteren (Beatty, 2003). De cognitieve interviewmethode kan tevens worden gebruikt om te onderzoeken of items of teksten op dezelfde manier geïnterpreteerd worden door respondenten met verschillende (culturele of etnische) achtergronden (bijv. McCoy, 2010; Warnecke et al., 1997). In het afgelopen decennium is een aantal crossculturele studies verricht waarin gebruik is gemaakt van de cognitieve interviewmethode om de conceptuele equivalentie van instrumenten in kaart te brengen (Agans, Deeb-Sossa & Kalsbeek, 2006; McCoy, 2010; Nápoles-Springer et al., 2006; Willis et al., 2008). Tot dusver hebben deze studies zich voornamelijk gericht op het pretesten van vragenlijsten die gebruikt worden in de gezondheidszorg (bijv. McCoy, 2010; Nápoles-Springer et al., 2006). Zo hebben Nápoles-Springer en collega's (2006) het nut van cognitieve interviews aangetoond voor het pretesten van een instrument dat in de Verenigde Staten veel wordt gebruikt om de kwaliteit van de gezondheidszorg te evalueren. In maar liefst 126 van de totaal 159 items werden systematische etnische verschillen gevonden in de interpretatie van de items en het gebruik van de antwoordopties. Het item 'Hoe vaak keek de arts naar u terwijl u aan het praten was?' werd bijvoorbeeld niet op dezelfde manier geïnterpreteerd door de verschillende etnische groepen. Een Afro-Amerikaanse vrouw gaf aan dat ze zich erg ongemakkelijk zou voelen als een arts naar haar zou kijken. Voor Latijns-Amerikaanse deelnemers was het onduidelijk of bedoeld werd of de arts oogcontact maakt of dat hij naar het lichaam kijkt. Recent is in Nederland een eerste studie verricht waarin de cognitieve interviewmethode gebruikt is om inzicht te verkrijgen in de conceptuele equivalentie van een selectie-instrument, in dit geval een rollenspel (Oostrom & Born, submitted). Resultaten lieten systematische etnische verschillen zien in de interpretatie van bijna alle aspecten van de instructietekst, zoals de interpretatie van bepaalde woorden en zinnen, van het beschreven probleem en van de beschreven context.

Een aantal factoren kan de effectiviteit van de cognitieve interviewmethode echter beïnvloeden, zoals de selectie van respondenten, de typen vragen die gesteld worden, de selectie en training van de interviewers en de analysemethode. Hieronder zullen deze factoren verder worden uitgelegd.

3.1 De selectie van respondenten

De interviews vinden meestal plaats in een rustige, gecontroleerde omgeving bij een klein aantal doelgericht verworven respondenten (DeMaio, Rothgeb & Hess, 1998). Tijdens de interviews moeten de respondenten goed worden geïnstrueerd

over hun taak en dient te worden benadrukt dat het instrument getest wordt en niet de respondenten. De literatuur over de cognitieve interviewmethode verschaft echter weinig duidelijkheid over de ideale samenstelling en grootte van de steekproef van respondenten. Willis (2005) adviseert om bij de werving van respondenten te streven naar voldoende diversiteit (qua leeftijd, opleiding, geslacht en etniciteit) zodat alle mogelijke problemen die kunnen optreden bij het betreffende instrument ook daadwerkelijk naar voren komen. Als de cognitieve interviewmethode gebruikt wordt om te onderzoeken of items of teksten op dezelfde manier geïnterpreteerd worden door respondenten met verschillende etnische achtergronden, is het vanzelfsprekend van belang om een etnisch diverse groep respondenten te werven. Daarnaast is het volgens Willis van belang om respondenten te kiezen die representatief zijn met betrekking tot het doel en het gebruik van het instrument. Het heeft bijvoorbeeld geen zin om een persoonlijkheidsvragenlijst voor lageropgeleiden te pretesten onder een groep academische studenten. Doorgaans bestaan steekproeven uit 20 tot 30 respondenten (Willis & Miller, 2011). In de eerdere crossculturele studies waarin de cognitieve interviewmethode is toegepast, bestonden de steekproeven uit 16 (McCoy, 2010) tot 48 (Nápoles-Springer et al., 2006) respondenten, behorende tot verschillende etnische groepen.

3.2 De typen vragen

Er zijn twee typen cognitieve interviews te onderscheiden, namelijk (1) hardop-denken-interviews (*thinking aloud*) en (2) *probing* interviews (Collins, 2003). In beide soorten interviews vraagt de interviewer naar de gedachten van respondenten tijdens het lezen van teksten of items. De twee typen interviews verschillen echter wat betreft vorm, veronderstellingen, en het soort informatie dat zij opleveren. De voorkeur voor hardop-denken-interviews of *probing* interviews hangt onder meer af van de leeftijd en de cognitieve vaardigheden van de respondenten, de inhoud, het doel en de toepassing van het instrument.

Zoals de naam al aangeeft, vraagt de interviewer in de hardop-denken-methode de respondent om alle gedachten die bij hem of haar opkomen hardop uit te spreken. Interviewers hebben hierbij een passieve rol; zij stimuleren de respondent slechts om gedachten hardop te blijven uitspreken. Een voordeel van deze methode is dat de kans dat de interviewer het denkproces stuurt of beïnvloedt, erg klein is (Oksenberg, Cannell & Kalton, 1991). Volgens sommigen is een tweede voordeel dat interviewers geen specifieke kennis hoeven te hebben van testontwikkeling of het instrument in kwestie (Bolton & Bronkhorst, 1996). Een nadeel van deze methode is echter dat het een behoorlijke verbale vaardigheid vraagt van de respondent en bovendien kan het hardop denken de cognitieve processen verstoren. De respondent kan bijvoorbeeld worden afgeleid van de taak of de taak op een bewustere manier gaan vervullen (Willis, 2005).

In *probing* interviews heeft de interviewer specifieke vragen (*probes*) voorbereid. Een voorbeeld van een *probe* is 'Hoe heeft u de term *X* geïnterpreteerd?'. Verschillende auteurs zijn van mening dat *probing* interviews voordelen bieden ten opzichte van hardop-denken-interviews. Willis (2005) is van mening dat door *probing* het gedrag van de respondenten beter gestuurd kan worden. Respondenten

kunnen soms helemaal opgaan in irrelevante gedachten. Door specifieke vragen te stellen kan de interviewer de respondent 'bij de les houden'. Vanzelfsprekend moet de interviewer wel kennis hebben van wat de relevante aspecten (bijv. mogelijke typen bias) voor de betreffende teksten of items zijn en met welke vragen de aandacht van respondenten op deze aspecten kan worden gericht. Een tweede voordeel van *probing* is dat het informatie oplevert die mogelijk niet naar voren zou komen als de interviewer hier niet specifiek naar vraagt (Willis, 2005). In hardop-denken-interviews is het onzeker of de informatie die naar voren komt, nuttig zal zijn met betrekking tot het oplossen van problemen in het instrument (bijv. problemen met betrekking tot de interpretatie van een bepaald woord), terwijl in *probing* interviews direct naar deze problemen kan worden gevraagd. Een nadeel van *probing* is wellicht dat de interviewer de inhoud en het verloop van het interview stuurt. Oksenberg en collega's (1991) raden daarom aan om niet te veel probes te gebruiken na elk item of tekstgedeelte. Daarnaast is de kwaliteit van *probing* interviews geheel afhankelijk van de kwaliteit van de probes. Het is dus van belang om veel aandacht te schenken aan de ontwikkeling van de *probes*. Het onderzoeksteam en de interviewer hebben hier allebei een belangrijk aandeel in (Beatty & Willis, 2007). Goede *probes* moeten niet een 'juist' antwoord doen vermoeden (Willis, 2005) en moeten naar specifieke onderdelen of gedachten vragen in plaats van naar algemene ervaringen (Foddy, 1998).

3.3 De selectie en training van interviewers

Interviewers kunnen twee verschillende rollen hebben, namelijk die van dataverzamelaar en/of van onderzoeker. Volgens Conrad en Blair (2004) zouden interviewers vooral de rol van dataverzamelaar moeten hebben en bepalen onderzoekers de *probes* die zij gebruiken tijdens de interviews. Dit vergt vanzelfsprekend een zeer grondige voorbereiding. Een alternatief is dat interviewers zelf de rol van onderzoeker op zich nemen en de vrijheid hebben om tijdens het interview *probes* te gebruiken die op dat moment relevant lijken (Beatty & Willis, 2007). Dergelijke interviews zijn van meer exploratieve aard. Een voordeel is dat de interviewer problemen kan aansnijden die gemist kunnen zijn in het voorbereidingsproces. Vanzelfsprekend moet de interviewer in het laatste type interview wel in staat zijn om deze problemen op te merken en ter plekke geschikte *probes* kunnen bedenken. Willis (2005) beschrijft een trainingsprogramma waarin vragenlijst-ontwerp, observaties van de interviews, soorten *probes*, valkuilen en andere praktische zaken worden behandeld. Een adequate training lijkt onontbeerlijk, want een gebrek aan kennis en vaardigheden brengt de kwaliteit van de cognitieve interviews in gevaar. Dit kan op zijn beurt een negatieve invloed hebben op bijvoorbeeld de betrouwbaarheid en de validiteit van de cognitieve interviewmethode. Er kan een parallel worden getrokken met de recent geïntroduceerde rollenspelers training in Assessment Centers (AC's). Schollaert en Lievens (2010) bespreken een rollenspelers training waarin het gebruik van prompts centraal staan. Prompts kunnen gedefinieerd worden als 'vooraf bepaalde en aangeleerde verbale en non-verbale handelingen tijdens de AC-opdrachten, die door rollenspelers over alle kandidaten heen consistent gehanteerd worden, met als doel dimensiegerelateerd gedrag

uit te lokken' (Schollaert & Lievens, 2010, p. 214). Deze prompts zijn tot op zekere hoogte vergelijkbaar met het gebruik van probes tijdens de cognitieve interviews. Met de prompts wordt echter dimensiegerelateerd gedrag uitgelokt, terwijl met de probes naar gedachten gevraagd wordt die opkomen bij het lezen van een tekst of item.

Het aantal interviewers is afhankelijk van praktische aspecten, zoals het totaal aantal te voeren interviews. In de eerdere crossculturele studies waarin de cognitieve interviewmethode is toegepast, werd gebruikgemaakt van één getrainde interviewer (Oostrom & Born, submitted) tot vier getrainde interviewers (McCoy, 2010; Nápoles-Springer et al., 2006). Kudela et al. (2004) lieten in een crossculturele studie zien dat bij het pretesten van een vragenlijst dezelfde problemen met betrekking tot het begrijpen en interpreteren van de items naar voren kwamen ongeacht het team van interviewers en hun etniciteit.

3.4 *De manieren om de interviewgegevens te analyseren*

De cognitieve interviewmethode levert uiteindelijk een grote hoeveelheid informatie op die geëvalueerd moet worden om te bepalen of de teksten en items in het instrument (bijv. de persoonlijkheidsvragenlijst, de interessevragenlijst) wel of niet op de juiste manier geïnterpreteerd worden. Er zijn drie analysetechnieken: (1) kwalitatief, (2) kwantitatief, door middel van gedragscodering, en (3) een combinatie van kwalitatief en kwantitatief. Deze drie technieken zullen hieronder worden besproken.

Bij kwalitatieve analyses verstrekt de interviewer aantekeningen van zijn of haar observaties met betrekking tot de manier waarop het item of de tekst geïnterpreteerd wordt. Deze aantekeningen worden na afloop van de interviews geanalyseerd door de onderzoeker of het onderzoeksteam (Miller, Mont, Maitland, Altman & Madans, 2010). Een voorbeeld van een kwalitatieve analyse techniek is 'opeenvolgende aggregatie', waarbij elke interviewer eerst per item de resultaten van alle interviews samenvat. Vervolgens combineert de onderzoeker de resultaten van de verschillende interviewers, wat resulteert in een samenvatting van observaties voor ieder item (Willis, 2005). De interviews waarbij interviewers de vrijheid hebben om spontane *probes* te gebruiken, maken de data-analyse complexer. De inhoud van deze interviews kan namelijk aanzienlijk van elkaar verschillen. Daarnaast zal in elk interview de nadruk komen te liggen op andere problemen.

Traditioneel hing de evaluatie van de items of teksten vooral af van de percepties van de interviewer tijdens de interviews (Fowler & Cannell, 1996). Recentelijk wordt er bij het analyseren van de interviews echter steeds meer gebruikgemaakt van kwantitatieve technieken, zoals systematische gedragscodering. Gedragscodering betreft een methode waarbij de interviews worden opgenomen met een camera of recorder. Het gedrag van de respondenten wordt achteraf door de onderzoekers of andere experts geobserveerd en gecodeerd (DeMaio et al., 1998). Deze codering levert systematische en repliceerbare resultaten op die gebruikt kunnen worden om de bron van het probleem in het item of de tekst te lokaliseren, bijvoorbeeld een misinterpretatie van een bepaald woord (Fowler, 1995). De frequentie waarmee bepaalde problemen optreden, wordt bijgehouden. Items of

tekstgedeelten die regelmatig problemen opleveren, bijvoorbeeld in 15% van de interviews, worden vervolgens opnieuw bekeken door de onderzoekers en verwijderd of aangepast. Een nadeel van deze methode is echter dat het geen inzicht geeft in waarom deze items of tekstgedeelten problemen opleveren. Blair, Ackermann, Piccinino en Levenstein (2006) raden daarom aan om gedragscodering te combineren met een kwalitatieve analyse van de antwoorden van de respondenten. Op basis van de kwalitatieve analyse wordt bekeken hoe items of teksten dienen te worden gereviseerd. Oostrom en Born (submitted) hebben deze methode van data-analyse toegepast. Uit de gedragscodering bleken systematische etnische verschillen in de interpretatie van bijna alle aspecten van de instructieteksten van de rollenspellen. Uit de kwalitatieve analyse van de data bleek onder meer culturele verschillen in attributiestijlen een rol te spelen. Ter illustratie: in de instructietekst werd beschreven dat een werkgroep niet goed functioneerde. Leden van de etnische minderheidsgroep attribueerden het slechte functioneren van de werkgroep aan de leider van de werkgroep. Autochtone deelnemers attribueerden hetzelfde gedrag aan de capaciteiten van de leden van de werkgroep. Dit probleem zou kunnen worden opgelost door in de instructietekst de oorzaken van het gedrag van de werkgroep expliciet te beschrijven.

De resultaten van Oostrom en Born (submitted) doen vermoeden dat de cognitieve interviewmethode zowel 'rechtvaardige' als 'onrechtvaardige' etnische scoreverschillen aan het licht brengt. Met 'rechtvaardige' scoreverschillen worden scoreverschillen bedoeld die overeenkomen met de verschillen in het concept waarop de etnische groepen worden vergeleken en die relevant zijn voor de uitoefening van de toekomstige functie. Het is echter een lastige opgave om 'rechtvaardige' scoreverschillen te onderscheiden van 'onrechtvaardige' scoreverschillen, aangezien er dan metingen van het concept beschikbaar dienen te zijn die vrij zijn van elke vorm van testbias.

4 Kritische reflectie op de cognitieve interviewmethode

De cognitieve interviewmethode is een relatief nieuwe, kwalitatieve methode om testbias te onderzoeken en te reduceren. Het is een op het instrument gerichte strategie voor het inperken van etnische scoreverschillen, die overeenkomsten heeft met het minimaliseren van de benodigde taalkennis, het gebruiken van items die even (on)bekend zijn voor verschillende etnische groepen, statistische procedures (zoals DIF) en het gebruiken van review panels (strategieën 12, 13, 14 en 15; zie tabel 1). Zowel de cognitieve interviewmethode als het minimaliseren van de benodigde taalkennis (strategie 12) heeft als doel etnische scoreverschillen die door taalgebruik veroorzaakt worden te reduceren. De cognitieve interviewmethode is echter gericht op het creëren van conceptuele equivalentie, terwijl het minimaliseren van de benodigde taalkennis gericht is op het beperken van verbale informatie. Items die even (on)bekend zijn voor verschillende etnische groepen (strategie 13), zijn meestal zogeheten 'inhoudsvrije' items. Het blijkt echter voor veel instrumenten onmogelijk te zijn om 'inhoudsvrije' items of teksten te ontwikkelen (Poortinga, 2001). De cognitieve interviewmethode lijkt dan een

geschiktere, breder toepasbare methode om etnische scoreverschillen te minimaliseren dan het gebruiken van items die even (on)bekend zijn voor verschillende etnische groepen. Statistische procedures, zoals DIF (strategie 14), en de cognitieve interviewmethode zouden theoretisch gezien dezelfde items moeten detecteren. De cognitieve interviewmethode kan echter ook worden gebruikt om de interpretatie van instructies, antwoordschalen en open vragen te pretesten. Daarnaast verschaft de cognitieve interviewmethode meer informatie over de oorzaak van het probleem en handvatten voor het verbeteren van de teksten en items. Zowel bij de cognitieve interviewmethode als bij review panels (strategie 15) wordt gebruikgemaakt van doelgericht verworven respondenten die beoordelen of teksten of items gebruikt kunnen worden bij het testen van kandidaten met verschillende etnische achtergronden. De cognitieve interviewmethode maakt echter gebruik van respondenten die representatief zijn voor de groep waarbij het instrument wordt afgenomen en onderzoekt of items of teksten op dezelfde manier geïnterpreteerd worden door respondenten met verschillende etnische achtergronden. Review panels bestaan daarentegen uit individuen die kennis hebben van culturele verschillen en die teksten en items voornamelijk beoordelen op het gebruik van stereotypen of kwetsend taalgebruik (bijv. Reckase, 1996).

De zoektocht naar effectieve strategieën die geen afbreuk doen aan de voorspelende waarde van het instrument, blijft continueren. In het huidige artikel wordt de cognitieve interviewmethode beschreven als een mogelijke manier om onrechtvaardige scoreverschillen van diverse soorten selectie-instrumenten in te perken. Zoals in paragraaf 3 is beschreven, is de cognitieve interviewmethode reeds onderzocht op wetenschappelijke waarde, bruikbaarheid, en nuttigheid in de gezondheidszorg (bijv. McCoy, 2010; Willis & Miller, 2011). Er zijn echter enkele uitdagingen en suggesties voor verder onderzoek binnen de personeelsselectie.

Allereerst kunnen er problemen zijn met betrekking tot de betrouwbaarheid van de gegevens die de cognitieve interviewmethode oplevert. Een andere groep onderzoekers of interviewers zou tot andere conclusies kunnen komen na het cognitief pretesten van eenzelfde instrument (Forsyth, Rothgeb & Williams, 2004). Verschillen kunnen optreden door bijvoorbeeld het gebruik van verschillende steekproeven. Voor het onderzoeken van conceptuele equivalentie is het belangrijk om een gemixte steekproef te trekken qua etnische achtergrond. De vraag is echter hoever je daarin als onderzoeker moet gaan. In de selectiepraktijk wordt vaak onderscheid gemaakt tussen de autochtone en allochtone groep. De allochtone groep betreft echter een groep mensen met zeer diverse achtergronden, zelfs wanneer binnen het land van herkomst wordt gekeken (Altink, 2012). Op welke groepen richt je je als onderzoeker? Het lijkt niet de bedoeling om idiosyncratische antwoorden te krijgen als een instrument breed ingezet wordt in de selectiepraktijk. De literatuur over de cognitieve interviewmethode verschaft nog weinig duidelijkheid over de ideale samenstelling en grootte van de steekproef van respondenten.

Ten tweede dienen onderzoekers ervoor zorg te dragen dat de analyses tot inhoudsvalide conclusies leiden en niet incompleet of misleidend zijn (Beatty & Willis, 2007). De cognitieve interviews kunnen problemen identificeren die in de

selectiepraktijk niet zouden optreden. Studies naar de generaliseerbaarheid van de cognitieve interviewmethode zijn schaars, maar hebben wel bewijs gevonden dat de problemen die naar aanleiding van de interviews geïdentificeerd waren, ook in de praktijk een rol speelden (Beatty, Fowler & Cosenza, 2006; Willis & Schechter, 1997). Omdat de interviews in een gecontroleerde omgeving bij een klein aantal respondenten plaatsvinden, is het ook mogelijk dat niet alle eventuele problemen die in de praktijk een rol kunnen spelen, ook naar voren komen tijdens de interviews. Een assumptie van de cognitieve interviewmethode is dat de belangrijkste problemen in een vragenlijst of tekst onder elke groep respondenten die redelijk overeenkomt met de doelgroep van het instrument, zullen worden ontdekt (Beatty & Willis, 2007). De mogelijkheid blijft echter altijd bestaan dat een extra interview of een andere interviewer weer andere inzichten kan opleveren.

Ten derde lijkt de cognitieve interviewmethode meer geschikt voor bepaalde selectie-instrumenten dan voor andere. Het pretesten van selectie-interviews kan bijvoorbeeld problemen opleveren omdat de respondent en de cognitieve interviewer moeten schakelen tussen de rol van respondent/interviewer voor het selectie-interview en die van respondent/interviewer voor het cognitieve interview (zie ook Collins, 2003). Daarnaast is het te verwachten dat aanpassingen naar aanleiding van de uitkomsten van de cognitieve interviewmethode de grootste effecten voortbrengen bij selectie-instrumenten waarbij het juist interpreteren van de items een directe doorwerking heeft in de score, zoals bij cognitieve vaardigheidstests en persoonlijkheidsvragenlijsten. Oostrom en Born (submitted) toonden aan dat ook instructieteksten van rollenspelen anders geïnterpreteerd kunnen worden door de etnische meerderheidsgroep en de etnische minderheidsgroep. De vraag blijft echter in hoeverre deze verschillen in interpretaties doorwerken in de uiteindelijke scores op het rollenspel.

Het bovenstaande leidt tot twee fundamentele vragen. De eerste vraag luidt of het aanpassen van de items of teksten op basis van de gegevens die de cognitieve interviewmethode oplevert, daadwerkelijk tot statistisch significant minder etnische scoreverschillen leidt. De tweede vraag luidt of het aanpassen van de items of teksten op basis van de gegevens die de cognitieve interviewmethode oplevert, de predictieve validiteit van het selectie-instrument niet doet dalen. De tot op heden gepubliceerde artikelen over de cognitieve interviewmethode zijn gericht geweest op het achterhalen van problematische items en teksten. Zover bekend, zijn er nog geen artikelen gepubliceerd waarin de daadwerkelijke gevolgen van de aanpassingen van de items of teksten voor de etnische scoreverschillen en de validiteit van het instrument geïllustreerd worden. In tegenstelling tot andere strategieën om etnische scoreverschillen te reduceren die enkel de problematische items detecteren, geeft de cognitieve interviewmethode inzicht in de onderliggende oorzaken en handvatten voor het aanpassen van teksten en items. Wij verwachten dat het aanpassen van de teksten en items leidt tot kleinere etnische scoreverschillen, omdat hierdoor systematische irrelevante testvariantie binnen de etnische minderheidsgroep gereduceerd wordt. Tevens verwachten wij dat het verwijderen van systematische irrelevante testvariantie zorgt voor een meer zuivere meting van het predictor construct en dat de voorspellende waarde van het

instrument hierdoor zelfs zou kunnen toenemen. De cognitieve interviewmethode zou daarom een waardevolle methode kunnen zijn om etnische scoreverschillen te reduceren. Er dient echter nog onderzoek te worden gedaan naar de effectiviteit van deze strategie met betrekking tot het daadwerkelijk reduceren van etnische scoreverschillen, de gevolgen voor de validiteit van het selectie-instrument en ook de utiliteit van deze strategie.

5 Conclusie

De samenstelling van de beroepsactieve bevolking wordt steeds diverser. Het blijkt een complexe uitdaging te zijn voor organisaties om deze diversiteit te weerspiegelen in hun personeelsbestand. Selectie-instrumenten zijn namelijk veelal primair ontwikkeld en gevalideerd voor de autochtone bevolkingsgroep (Derous et al., 2011) en genereren dikwijls etnische scoreverschillen (Ployhart & Holtz, 2008). Aangezien de omvang van de scoreverschillen bepaalt of een selectie-instrument 'adverse impact' vertoont, moet bij de ontwikkeling van selectie-instrumenten getracht worden om scoreverschillen tussen de autochtone groep en de etnische minderheidsgroep zo veel mogelijk te reduceren. Er zijn reeds veel strategieën ontwikkeld om etnische scoreverschillen op selectie-instrumenten te verminderen, maar deze blijken niet altijd effectief en doen vaak afbreuk aan de voorspellende waarde van het selectie-instrument (Ployhart & Holtz, 2008). In het huidige artikel werd een methode geïntroduceerd die onderzoekers en testontwikkelaars kan helpen in het reduceren van onrechtvaardige scoreverschillen tussen etnische groepen, namelijk de cognitieve interviewmethode. Deze methode kan gebruikt worden om te onderzoeken of items of teksten inhoudelijk dezelfde betekenis toegeschreven krijgen door respondenten met verschillende etnische achtergronden. De interviews leveren kwalitatieve informatie op die bij een kleine groep respondenten verzameld is. De interviews kunnen rijke informatie opleveren over het beantwoordingsproces en de fouten die hierin gemaakt worden, en geven daardoor inzicht in onderliggende oorzaken voor etnische verschillen in de interpretatie van teksten en items. Tot dusver is er – voor zover bij ons bekend is – slechts een studie verricht waarin de cognitieve interviewmethode gebruikt is om inzicht te verkrijgen in etnische scoreverschillen op een selectie-instrument, in dit geval een rollenspel (Oostrom & Born, submitted). De resultaten van deze studie tonen het nut van de cognitieve interviewmethode aan. De cognitieve interviewmethode zou dus een waardevol hulpmiddel kunnen zijn voor onderzoekers en testontwikkelaars in de ontwikkeling en verbetering van selectie-instrumenten. Door het achterhalen van de oorzaken van de etnische verschillen kunnen problemen namelijk gerichter aangepakt worden. Er zijn echter enkele uitdagingen en suggesties voor verder onderzoek met betrekking tot het gebruik van de cognitieve interviewmethode binnen de personeelsselectie. Fundamentele vragen met betrekking tot de samenstelling en grootte van de steekproef, de validiteit en de utiliteit van de cognitieve interviewmethode dienen nog beantwoord te worden. Wij hopen dat dit artikel het onderzoek naar de toe-

passing van de cognitieve interviewmethode in de ontwikkeling van selectie-instrumenten stimuleert.

Praktijkbox

Wat betekenen de resultaten voor de praktijk?

- Aangezien selectie-instrumenten primair ontwikkeld en gevalideerd zijn voor de autochtone bevolkingsgroep en dikwijls scoreverschillen tussen de autochtone groep en etnische minderheidsgroepen genereren, is het selecteren van een divers personeelsbestand voor veel organisaties een complexe uitdaging. In het huidige artikel is een methode geïntroduceerd die onderzoekers en testontwikkelaars hierbij kan helpen, namelijk de cognitieve interviewmethode. De cognitieve interviewmethode is een kwalitatieve methode die inzicht biedt in de cognitieve processen van kandidaten die in werking treden bij het beantwoorden van items of het lezen van teksten.
- De cognitieve interviewmethode kan gebruikt worden om te onderzoeken of items of teksten inhoudelijk dezelfde betekenis toegeschreven krijgen door respondenten met verschillende etnische achtergronden.
- Hoewel de cognitieve interviewmethode een waardevolle strategie lijkt te zijn voor onderzoekers en testontwikkelaars in de ontwikkeling en verbetering van selectie-instrumenten, dient er ook onderzoek te worden gedaan naar de effectiviteit van deze strategie met betrekking tot het daadwerkelijk reduceren van etnische scoreverschillen, de gevolgen voor de validiteit van het selectie-instrument en ook de utiliteit van deze strategie.

Literatuur

- Agans, R.P., Deeb-Sossa, N. & Kalsbeek, W.D. (2006). Mexican immigrants and the use of cognitive assessment techniques in questionnaire development. *Hispanic Journal of Behavioral Sciences*, 28, 209-230.
- Aguinis, H. (2004). *Test score banding in human resource selection: Legal, technical, and societal issues*. Westport, CT: Quorum.
- Altink, W. (2012). Wie of wat bepaalt hoe we groepen definiëren? Een belangrijke vraag voor de selectiepsychologie. *Gedrag & Organisatie*, 25, 275-284.
- Arthur, W., Edwards, B.D. & Barrett, G.V. (2002). Multiple-choice and constructed response tests of ability: Race-based subgroup performance differences on alternative paper-and-pencil test formats. *Personnel Psychology*, 55, 985-1008.
- Barofsky, I. (2000). The role of cognitive equivalence in studies of health-related quality-of-life assessments. *Medical Care*, 38, II-125-II-129.
- Beatty, P.C. (2003). Answerable questions: Advances in the methodology for identifying and resolving questionnaire problems in survey research. *Dissertation Abstracts International*, 64, 3504A.
- Beatty, P.C., Fowler, F.J. & Cosenza, C. (2006). Do questionnaire design recommendations lead to measurable improvements? Some experiments with alternate versions of complex survey questions. *Proceedings of the Q2006 European Conference on Quality in Survey Statistics*.

- Beatty, P.C. & Willis G.B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71, 287-311.
- Berry, C., Gruys, M. & Sackett, P.R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, 91, 696-705.
- Blair, J., Ackermann, A., Piccinino, L. & Levenstein, R. (2006). Using behavior coding to validate cognitive interview findings. *Proceedings of the American Statistical Association: Survey Research Methods Section*, 3896-3900.
- Bobko, P., Roth, P.L. & Potosky, D. (1999). Derivation and implications of a meta-analysis matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, 52, 561-589.
- Bolton, R.N. & Bronkhorst, T.M. (1996). Questionnaire pretesting: Computer assisted coding of concurrent protocols. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 37-64). San Francisco, CA: Jossey-Bass.
- Born, M.Ph. (2010). Het beoordelen van mensen voor werkgerelateerde doelen: Etniciteit, moraliteit, en subjectiviteit. *Gedrag & Organisatie*, 23, 232-256.
- Campion, M.A., Outtz, J.L., Zedeck, S., Schmidt, F.L., Kehoe, J.F., Murphy, K.R. et al. (2006). The controversy of score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology*, 54, 149-185.
- Ceci, S.J. & Williams, W.M. (1997). Schooling, intelligence, and income. *American Psychologists*, 52, 1051-1058.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.
- Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12, 229-238.
- Conrad, F.G. & Blair, J. (2004). Data quality in cognitive interviews: The case of verbal reports. In S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for determining cognitive and communicative processes in survey research* (pp. 177-195). San Francisco, SA: Jossey-Bass.
- De Corte, W. & Lievens, F. (2003). A practical procedure to estimate the quality and the adverse impact of single-stage selection decisions. *International Journal of Selection and Assessment*, 11, 89-97.
- De Corte, W., Lievens, F. & Sackett, P.R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*, 92, 1380-1393.
- DeMaio, T.J., Rothgeb, J. & Hess, J. (1998) Improving survey quality through pretesting. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 3, 50-58.
- Derous, E., Van der Velde, M. & Born, M.Ph. (2011). Terug van nooit weggeweest: Inleiding op de nieuwe reeks 'Werving en Selectie'. *Gedrag & Organisatie*, 24, 3-17.
- DeShon, R.P., Smith, M.R., Chan, D. & Schmitt, N. (1998). Can racial differences in cognitive test performance be reduced by presenting problems in a social context? *Journal of Applied Psychology*, 83, 438-451.
- Foddy, W. (1998). An empirical evaluation of in-depth probes used to pretest survey questions. *Sociological Methods and Research*, 27, 103-133.
- Forsyth, B., Rothgeb, J.M. & Williams, G.B. (2004). Does questionnaire pretesting make a difference? An empirical test using a field survey experiment. In S. Presser, J.M. Roth-

- geb, M.C. Couper, J.T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluation survey questionnaires*. Hoboken, NJ: John Wiley.
- Fowler, F.J. (1995). *Improving survey questions: Design and evaluation*. Applied social research methods series. Thousand Oakes, CA: Sage.
- Fowler, F.J. & Cannell, C.F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 15-36). San Francisco, CA: Jossey-Bass.
- Gittler, G. (1990). *The three-dimensional cube test 3DC. A Rasch-calibrated spatial ability test. Theoretical background and test manual*. Weinheim, Germany: Beltz Test.
- Grill, J.J. & Bariel, N.R. (1977). Language bias in tests: ITPA grammatic closure. *Journal of Learning Disabilities*, 10, 229-235.
- Gutjahr, G. (1985). *Psychology of interviews: Theory and practice*. Heidelberg, Germany: Sauer.
- Hamamura, T., Heine, S.J. & Paulhus, D.L. (2008). Cultural differences in response styles: The role of dialectical thinking. *Personality and Individual Differences*, 44, 932-942.
- Hattrup, K., Rock, J. & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology*, 82, 656-664.
- Hausknecht, J.P., Day, D.V. & Thomas, S.C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683.
- Holland, P.W. & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
- Hough, L.M., Oswald, F.L. & Ployhart, R.E. (2001). Determinants, detections and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9, 152-194.
- Hui, C.H. & Triandis, H.C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 296-309.
- Ironson, G.H., Guion, R.M. & Ostrander, M. (1982). Adverse impact from a psychometric perspective. *Journal of Applied Psychology*, 67, 419-432.
- Jensen, A.R. (1981). *Straight talk about mental tests*. New York, NY: Free Press.
- Kudela, M.S., Levin, K., Tseng, M., Hum, M., Lee, S., Wong, C. et al. (2004). *Tobacco use cessation supplement to the current population survey Chinese, Korean, and Vietnamese translations: Results of cognitive testing*. Rockville, MD: National Cancer Institute.
- Lievens, F. & Sackett, P.R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91, 1181-1188.
- McCoy, H. (2010). Using cognitive interviewing to explore causes for racial differences on the MAYSI-2. *Crime & Delinquency*. Advance online publication. doi: 10.1177/0011128710388922
- Miller, K., Mont, D., Maitland, A., Altman, B. & Madans, J. (2010). Results of a cross-national structured cognitive interviewing protocol to test measures of disability. *Quality & Quantity*, 4, 801-815.
- Nápoles-Springer, A.M., Santoyo-Olsson, J., O'Brien, H. & Stewart, A.L. (2006). Using cognitive interviews to develop surveys in diverse populations. *Medical Care*, 44, 21-30.
- Nguyen, N.T., McDaniel, M.A. & Whetzel, D. (2005, April). *Subgroup differences in situational judgment test performance: A meta-analysis*. Paper presented at the 20th annual conference of the Society for Industrial and Organizational Psychology (SIOP), Los Angeles, CA.
- Oksenberg, L., Cannell, C. & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics*, 7, 349-365.

- Oostrom, J.K. & Born, M.Ph. (submitted). Using cognitive pretesting to explore causes for ethnic differences on role-plays.
- Ployhart, R.E. & Holtz, B.C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153-172.
- Poortinga, Y.H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology*, 24, 737-756.
- Poortinga, Y.H. (2001). Het gebruik van psychologische tests bij allochtonen: Probleembeschrijving. In N. Bleichrodt & F.J.R. van de Vijver (Eds.), *Diagnostiek bij allochtonen: Mogelijkheden en beperkingen van psychologische tests* (p. 15-42). Lisse, Nederland: Swetz & Zeitlinger.
- Poortinga, Y.H. & Van de Vijver, F.J.R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology*, 18, 259-282.
- Reckase, M.D. (1996). Test construction in the 1990s: Recent approaches every psychologist should know. *Psychological Assessment*, 8, 354-359.
- Roth, P.L., Bevier, C.A., Bobko, P., Switzer, F.S., III & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297-330.
- Roth, P.L. & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology*, 85, 399-406.
- Rushton, J.P. (1995). *Race, evolution, and behavior: A life history perspective*. New Brunswick, NJ: Transaction.
- Ryan, A.M. (2001). Explaining the Black-White test score gap: The role of test perceptions. *Human Performance*, 14, 45-75.
- Ryan, A.M., Ployhart, R.E. & Friedel, L.A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology*, 83, 298-307.
- Ryan, A.M., Sacco, J.M., McFarland, L.A. & Kriska, S.D. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology*, 85, 163-179.
- Sackett, P.R. & Roth, L. (1991). A Monte Carlo examination of banding and rank order methods of test score use in personnel selection. *Human Performance*, 4, 279-295.
- Sackett, P.R., Schmitt, N., Ellingson, J.E. & Kabin, M.B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*, 56, 302-318.
- Schleicher, D.J., Van Iddekinge, C.H., Morgeson, F.P. & Campion, M.A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology*, 95, 603-617.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmitt, N. (1989). Fairness in employment selection. In M. Smith & I.T. Robertson (Eds.), *Advances in selection and assessment* (pp. 134-153). Chichester, England: Wiley.
- Schmitt, N., Clause, C.S. & Pulakos, E.D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. *International review of industrial and organizational psychology* (Vol. 11, pp. 115-139). New York, NY: Wiley.
- Schollaert, E. & Lievens, F. (2010). Een andere invalshoek op de rollenspelers training in Assessment Centers. *Gedrag & Organisatie*, 23, 213-231.

- Sin, H.P., Farr, J.L., Murphy, K.R. & Hausknecht, J.P. (2004). *An investigation of Black-White differences in self-selection and performance in repeated testing*. Paper presented at the 64th annual meeting of the Academy of Management. New Orleans, LA.
- Tam, A.P., Murphy, K.R. & Lyall, J.T. (2004). Can changes in differential dropout rates reduce adverse impact? A computer simulation study of a multi-wave selection system. *Personnel Psychology*, 57, 905-934.
- Tanzer, N.K., Gittler, G. & Ellis, B.B. (1995). Cross-cultural validation of items complexity in a LLTM-calibrated spatial ability test. *European Journal of Psychological Assessment*, 11, 170-183.
- Van de Vijver, F.J.R. & Tanzer, N.K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée*, 54, 119-135.
- Vodgel Matzen, L.B.L., Van der Molen, M.W. & Dudink, A.C.M. (1994). Error analysis of Raven test performance. *Personality and Individual Differences*, 16, 433-445.
- Warnecke, R.B., Johnson, T.P., Chávez, N., Sudman, S., O'Rourke, D.P., Lacey, L. & Horm, H. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, 7, 334-342.
- Whitney, D.J. & Schmitt, N. (1997). Relationships between culture and responses to bio-data employment items. *Journal of Applied Psychology*, 82, 113-129.
- Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.
- Willis, G.B., Lawrence, D., Hartman, A., Kudela, M., Levin, K. & Forsyth, B. (2008). Translation of a tobacco survey into Spanish and Asian languages: The tobacco use supplement to the current population survey. *Nicotine and Tobacco Research*, 10, 1075-1084.
- Willis, G.B. & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *Field Methods*, 23, 331-341.
- Willis, G.B. & Schechter, S. (1997). Evaluation of cognitive interviewing techniques: Do the results generalize to the field. *Bulletin de Méthodologie Sociologique*, 55, 40-66.

The cognitive interview method: Discussion of an alternative strategy for reducing test bias

Janneke K. Oostrom & Marise Ph. Born, Gedrag & Organisatie, volume 26, June 2013, nr. 2, pp. 205-222.

As selection instruments often generate ethnic score differences, creating a diverse work floor constitutes a complex challenge for many organizations. Therefore, in the development of selection instruments it should be aimed for to reduce unfair score differences between ethnic groups as much as possible. This article discusses a method that could help researchers and test developers in reducing ethnic score differences, namely the cognitive interview method. This qualitative method for evaluating questionnaires is already being used in the healthcare sector. In this article, it is argued that this method is useful in the field of personnel selection as well. The cognitive interview method can be used to study whether respondents with different ethnic background interpret items in a similar way.

Key words: cognitive interview method, cognitive pretesting, ethnic score differences, adverse impact, personnel selection